

SURVIVING THE VACUUM: A STRATEGY FOR SUSTAINING SOFTWARE IN THE ABSENCE OF RSE TEAMS

Stephan Druskat, Thomas Krause

Third Conference of Research Software Engineers, University of Birmingham,
Birmingham, UK, 4 September 2018.

Slides: <https://sdruskat.net/rse18-minimal-infrastructure/>



WHO WE ARE

- RSEs in corpus and theoretical linguistics respectively, at the Dept. of German Studies and Linguistics, Humboldt-Universität zu Berlin, Germany

Stephan Druskat

MA in English, Modern German Literature, Linguistics,

ORCID [0000-0003-4925-7248](#),

stephan.druskat@hu-berlin.de,  [@stdruskat](#),  [@sdruskat](#)

Thomas Krause

Computer Science Diploma,

ORCID [0000-0003-3731-2422](#), krauseto@hu-berlin.de,  [@thomaskrause](#)

"Emergency" strategy for research projects that produce software but don't have access to a centralized RSE team, to make their software sustainable/re-usable

WHAT WE DO

"Pharmakonzern"

```
(  
pos=/.FIN/ ->dep[func="sbj"]  
"Jugendliche"  
& cat="S" & #4 >secedge #3  
)
```

Q Search More History

3 matches
in 2 documents

Corpus List Search Options

Visible: All

Name	Texts	Tokens		
HSJ-Baseler_Maesebuch	10	10.020		
HSJ-Briefe	2	374		
HSJ-Maese_Nissima	4	4.867		
HSJ-Schevet_Jehude	9	11.547		
HSJ-Varia	11	22.918		
KAJUK	8	119.420		
KanDel_cross_cohort_v2	425	73.920		
KanDel_long_cohort1_v2	78	13.346		
KanDel_long_cohort2_v2	185	34.612		
kobaltL1v1.4	20	12.984		
kobaltL2v1.4	51	33.368		
Maerchenkorpus	211	295.880		
Mercurius	2	187.423		
MHD_context	4	2.760		
NoSta-D-1.4-bematac	22	25.934		
NoSta-D-1.4-falko	10	6.034		
NoSta-D-Anselm	2	2.710		
NoSta-D-Kafka	2	10.388		
NoSta-D-TueBaDZ	2	10.832		
NoSta-D-Unicum	2	11.312		
pcc2	2	399		
ridges.herbology	14	63.734		
RIDGES_Herbology_Versio	22	122.698		
RIDGES_Herbology_Versio	29	154.266		
RIDGES_Herbology_Versio	29	154.267		
Ridges_Herbology_Versio	13	60.811		
SMULTRON_Banana	2	3.782		

Help/Examples Query Result

Base text Token Annotations

Displaying Results 1 - 3 of 3 Result for: "Pharmakonzern" | (pos=/.FIN/ ->dep[func="sbj"]) ...

Path: pcc2 > 11299 (tokens 105 - 119)

left context: 5 right context: 5

dependencies (arches)

Information structure (grid)

Inf-Stat acc-gen giv-inactive

NP NP NP

PP PP

Sent s

Topic ab

tok) - darüber streiten , was Jugendliche wollen und brauchen , ohne auf die Idee

discourse referents (grid)

constituents (tree)

reference (discourse)

Feigenblatt Die Jugendlichen in Zossen wollen ein Musikcafé . Das forderten sie bei der ersten Zossener Runde am Dienstagabend . Dass die Politiker der Stadt dafür Verständnis haben , ist löblich . Mit dem Treffen im

The screenshot displays the **Annotation Graph Editor** interface, showing a hierarchical graph structure representing the sentence "The Rising was mounted by Irish republicans". The graph nodes are labeled N1 through N7, with their respective syntactic categories (S:NP, S:S, S:VP, S:PP, POS:DT, POS:VBG, POS:VBD, POS:VBN, POS:IN, POS:JJ, POS:NNS). Edges between nodes represent grammatical relations (D1-D12).

Sentence View:

- [0] The Easter Rising (Irish : Éirí Amach na Cásca)
- ☒ [1] The Rising was mounted by Irish republicans
- [2] It was the most significant uprising in Ireland
- [3] Organised by seven members of the Militant
- [4] Members of the Irish Volunteers - led by Sir John
- [5] There were isolated actions in other parts of the country
- [6] With vastly superior numbers and artillery
- [7] Most of the leaders were executed following the
- [8] Support for republicanism continued to grow
- [9] In December 1918, republicans (by then known as the IRA) began a campaign of
- [10] On 21 January 1919 they convened the First Dáil

Level View:

- # Level name
- ☐ 0 Sentences (detected)
- ☒ 1 POS
- ☒ 2 Syntax (Stanford)
- ☐ 3 NO ASSIGNED LEVEL *

Linked Sentences View:

Sentence [1] is linked to:

Console:

```
AtomicAL Console  
Working on corpus document.salt ("The Easter Rising (Irish: Éirí Amach na Cásca), a...").  
To display a list of available commands, type 'help'.  
p n1 n5 S:S  
help  
[command] Arguments Syntax example  
n (New structure node) [key]:[value] n pos:np s t1 t2 type:np  
s (New span node) [element][element][key]:[val] s t1 t2 type:np  
e (New edge) [-type] [source] [target] [key]:[value] e -d n1 n2 r:coref
```

WHAT IF WE'RE NOT THERE ANYMORE?
NO RSE TEAM ANYWHERE TO TAKE CARE OF SUSTAINABILITY

WE NEED A STRATEGY ALLOWING EVERYONE TO ...

- Create sustainable research software within project runtime**
 - "Side effects"**
 - Determine requirements for sustainable maintenance**
- Create evidence for making a case for installation of RSE teams**

STRATEGY

"MAKE DO"

- **Stick to best practices**
- **Run project as an open source project**

+ TWO BUILDING BLOCKS

BUILDING BLOCKS

- Engineer software for (some degree of) re-usability
 - Let infrastructure do the rest****

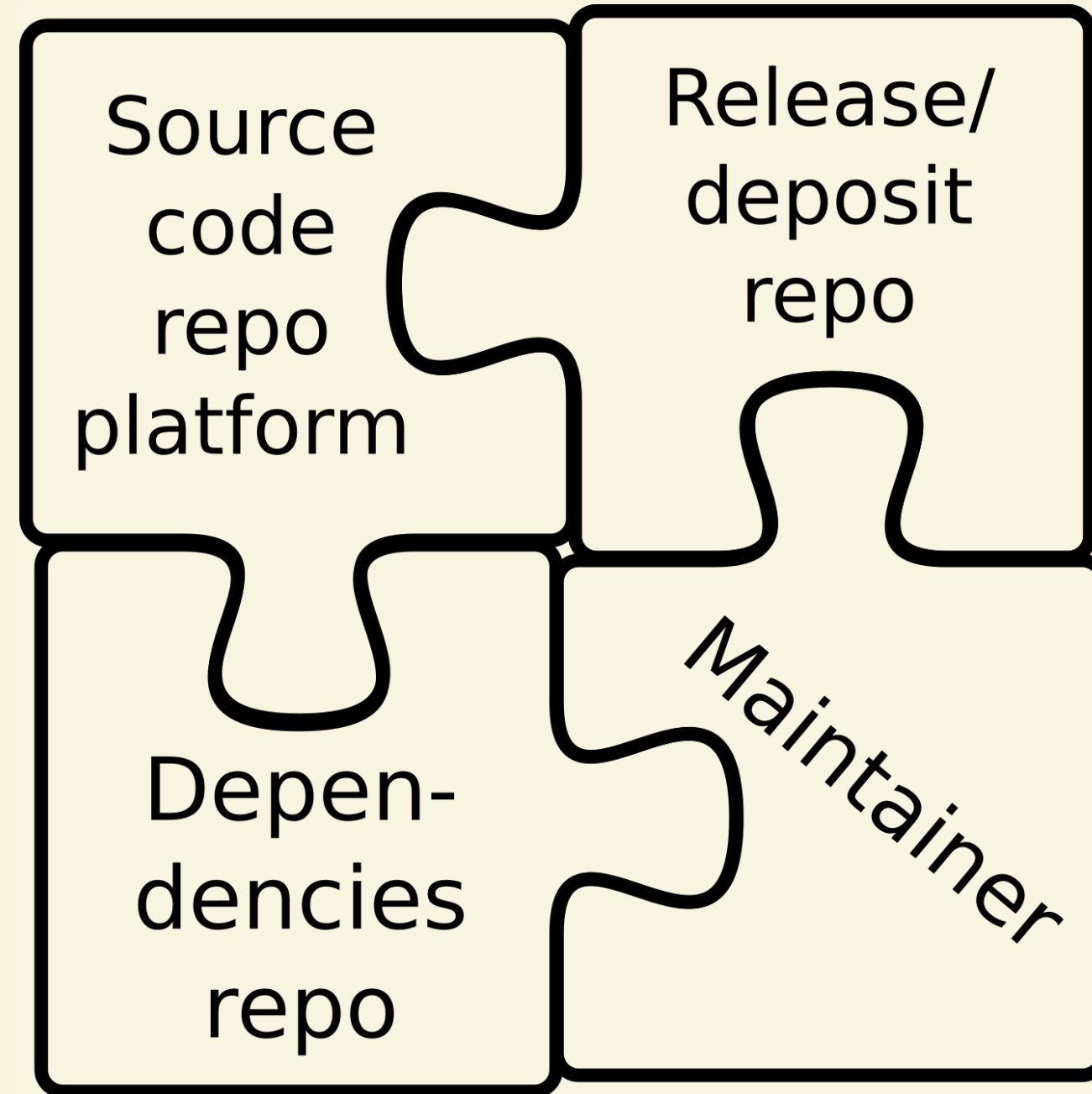
ENGINEER FOR RE-USABILITY

- 1. Use a generic data model**
- 2. Make software extensible by design**

INFRASTRUCTURE WHAT INFRASTRUCTURE?

- Use the free stuff that's there**
- Make sure you can exchange tools if necessary**
 - Implement an actual maintenance role**

MINIMAL INFRASTRUCTURE COMPONENTS



COMPONENT FUNCTIONS

Component	Function
Source code repo platform	Host code, docs, issues, landing page
Release & deposit repo	Long-term availability of artifacts, Version/citation metadata
Dependencies repo	Reproducible dependency graphs
Maintainer	Integrate, test, release, communicate, manage infrastructure, document/meta-document

"MAINTAINERS, MAINTAINERS, MAINTAINERS, ..."



- Incur cost, so hire (and train!) a student assistant**
- 10 hours / week should be enough in the long run**

OUR CASE

- Deep annotation of linguistic corpora
- Morph architectural & functional prototype
 - Generic data model (generic graph)
 - Extensible (Eclipse RCP/OSGi)

INFRASTRUCTURE

- GitHub**
- Zenodo**
- Maven Central, eclipse.org P2, P2 via GitHub**

INFRASTRUCTURE SUSTAINABILITY

- GitHub > Software Heritage
- Zenodo (long-term strategy)
- Dependency repos: system-critical, foundations,
Software Heritage
 - (Plan for hot-swapping)

MAINTAINERS

- Predetermined breaking point**
- 4 maintainership changes during project**

DOCUMENT! ALL! THE! THINGS!



CONCLUSION

- **Minimal requirements for technically sustainable software (?)**
- **Minimal infrastructure for sustainable software development (?)**

WHAT WILL WE HAVE LEARNED (IDEALLY)?

- Sustainable, re-usable software**
- Requirements for sustainable maintenance**
- Evidence for making a case for RSE teams!**

THANKS!

RSE18 Mentoring Programme: Neil Chue Hong

The Software Sustainability Institute for funding assistance (EPSRC, BBSRC, ESRC Grant EP/N006410/1)

Project funding: DFG grant GA 1288/11 ("Research software sustainability")

